Handbook of Research on Driving STEM Learning With Educational Technologies

María-Soledad Ramírez-Montoya
Tecnologico de Monterrey, Mexico

A volume in the Advances in Educational Technologies and Instructional Design (AETID) Book Series

www.igi-global.com

Published in the United States of America by
IGI Global
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA, USA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

Library of Congress Cataloging-in-Publication Data

CIP Data Pending
ISBN: 978-1-5225-2026-9
eISBN: 978-1-5225-2027-6

This book is published in the IGI Global book series Advances in Educational Technologies and Instructional Design (AETID) (ISSN: 2326-8905; eISSN: 2326-8913)

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

Chapter 4
Data Literacy and Citizenship: Understanding 'Big Data' to Boost Teaching and Learning in Science and Mathematics

Eddy L. Borges-Rey
University of Stirling, UK

ABSTRACT

This chapter explores the challenges that emerge from a narrow understanding of the principles underpinning Big data, framed in the context of the teaching and learning of Science and Mathematics. This study considers the materiality of computerised data and examines how notions of data access, data sampling, data sense-making and data collection are nowadays contested by datafied public and private bodies, hindering the capacity of citizens to effectively understand and make better use of the data they generate or engage with. The study offers insights from secondary and documentary research and its results suggest that understanding data in less constraining terms, namely: a) as capable of secondary agency, b) as the vital fluid of societal institutions, c) as gathered or accessed by new data brokers and through new technologies and techniques, and d) as mediated by the constant interplay between public and corporate spheres and philosophies, could greatly enhance the teaching and learning of Science and Mathematics in the framework of current efforts to advance data literacy.

INTRODUCTION

Notorious events such as the Flash Crash of 2:45 (Treanor, 2015) or Facebook's secret behavioural experiment on users (Meyer, 2014) have stirred global outcry. The first one, a 2010 stock market crash triggered by computerised algorithms, caused the largest intraday point decline in the entire history of the Dow Jones in a matter of minutes. The second one took place in 2012 and involved the manipulation of 689,003 users' feeds by ranking algorithms for a week without any kind of user consent. Unprecedentedly, both events led to two realisations: firstly, our knowledge of the obscure data technologies and agents mediating our everyday life is limited. And secondly, we are powerless in holding datafied bodies and technologies to account for this kind of actions.

DOI: 10.4018/978-1-5225-2026-9.ch004

As algorithms use data to make vital decisions about our lives in a domain free of public or civic scrutiny, practices such as surveillance, biometrics, automation, consumer profiling, algorithmic predictability, and machine learning tend to agitate public opinion. Simultaneously, in a dynamic indistinguishable to the public eye, mediated discourses of innovation extol Big data's messianic virtues as the cure to all societal illnesses, framing it as the ultimate panacea. Whilst reports on the marvels and failures of Big data populate the mainstream news agenda, citizens appear to be inadequately equipped to engage on equal terms with governments and corporations in the construction of a reality increasingly modelled by informational data. As numeracy tends to be rather limited across the board, a growing need for citizens to be able to understand the dynamics underpinning data is generally unfulfilled. Paradoxically, the idea of citizen empowerment through the use of ICTs remains a key objective in the era of Big data, primarily driven by expectations that new technologies and platforms will facilitate more responsive governments and provide people with access to information that will engender economic growth as well as creative and social fulfilment, especially after the launch of the Open Data Charter at the G8 summit in 2013.

Despite the efforts of the G8 governments to open up their data stores for public scrutiny, the techniques and strategies used to filter databases and datasets, identify and isolate noteworthy information from numerical data, and translate mathematical abstractions into insight that informs and reinforces decisions at the different levels of society, remain generally excluded from the education system. The teaching of Science and Mathematics appears to favour instrumental aspects of the discipline (Chevallard, 2013) to the detriment of the critical and fundamental approaches to Big data that are essential for the integral education of individuals living in this increasingly data-centric society.

This chapter thus aims to make a contribution to existing debates surrounding Science and Mathematics education by highlighting a set of challenges that emerge when Big data is not properly contextualised in the delivery of educational strategies for the enhancement of data literacy. In this vein, this study considers firstly the materiality of computerised data to examine its implications for data literacy. And secondly, it examines how notions of data access, data sampling, data sense-making and data collection are nowadays intermediated or contested by datafied governments and corporations, hindering the capacity of citizens to effectively understand and make better use of the data they generate or engage with. This article contributes to ongoing discussions on data literacy (Schield, 2004; Mandinach & Gummer, 2013; Stephenson & Schifter Caravello, 2007; Hunt, 2004; Qin & D'Ignazio, 2010; and Jacobs et al, 2009) and big/open data and algorithms as culture (Manovich, 2001; Pavlik, 2001; Howard, 2012; Hayles, 2012; boyd & Crawford, 2012; Mayer-Schonberger & Cukier, 2013; Manovich, 2013; O'Reilly, 2013; Crawford et al., 2014; Kitchin, 2014; Puschmann & Burgess, 2014; Struijs et al., 2014) in order to explore how data literacy is mediated by notions of Big data.

Quantifying the World

The superabundance of data in contemporary developed societies has produced a pressing necessity to understand the intricacies of Big data and harness its potential. boyd and Crawford (2012) remark that current efforts to conceptualise this socio-technical phenomenon elicit both utopic and dystopic rhetoric, whereby Big data is either constructed as a powerful solution to many societal issues, or a neoliberal force which threatens to constrain civic freedom by reducing citizens' right to privacy. More seriously, explain the authors, "the currents of hope and fear often obscure the more nuanced and subtle shifts that are underway" (boyd and Crawford, p. 663-4). With the salient emergence of Big data, the different social actors responsible for the categorisation of reality and the institution of knowledge have had

to rethink their ways of interacting with information (boyd & Crawford, 2012). Through this process, scholars, innovators and technologists have used Big data to engage in the quantification of the world in order to translate its complexities into manageable computing variables (Mayer-Schonberger & Cukier, 2013; Crawford, et al., 2014).

Statistics and Knowledge Production

The role of statistics in an increasingly number-centric society is to quantify relevant aspects of social life to objectively interpret its peculiarities. This information is then used in public deliberation, policy-making, business decisions, scientific research, education, etc. (Struijs et al., 2014). Mattelart suggests that a "social project inspired by a blind belief in numbers" and the establishment of mathematics as the imperative model for reasoning were instituted during the seventeenth and eighteenth centuries (2005, p. 5). Social research started to shift from a strongly theoretical attitude towards a more quantitative one, and during the nineteenth century "statistics –numeric statements about social life– became an authoritative way to describe social problems" (Best, 2012, p. 12-3). In this respect, Mattelart explains that countable measurements became "the prototype for truthful discourse" and spurred a quest for "the perfectibility of human society" (2005, p. 5). Knowledge-production institutions started to use statistics as numeric 'hard facts' (Best, 2012) and citizens appeared to ignore the fact that numerical figures are social constructs mediated by social productive forces and their accounts cannot be uncritically treated as being unreservedly true or factual (Best, 2012).

This reaction from the public is arguably a result of the low levels of numeracy across the population; and institutions responsible for the production of knowledge, such as news media, schools, universities, governments, etc., have operated unchallenged and free from civic scrutiny for years. Scholarly research on the use of statistics by news organisations, for instance, has sometimes been strongly critical and appears to agree that the report of statistical figures often shows signs of data misuse and mathematical inaccuracy (Maier, 2002; Cohn & Cope, 2012; Moore, 2009; Huff, 1973; Best, 2012). Maier (2002), on this subject, categorises a number of errors that frequently appear in news media, such as misinterpretation of numbers, misuse of mathematical terminology, inappropriate baseline, story–chart inconsistency, needless numbers, meaningless precision, or sensationalised numbers, amongst others. Likewise, Moore claims that external pollsters systematically misled audiences about the accuracy of their polls, "claiming a degree of precision in assessing public opinion that [was] far removed from reality" (2009, p. X). Therefore, it seems clear that the rhetoric of statistics –solidly embedded in a fact-centric culture– which tends to sensationalise, inflate, confuse, and oversimplify (Huff, 1973) is likely to affect the perceived legitimacy of figures and numbers in the public's eye (Best, 2012). Although the role of statistics during this period of data prevalence is yet to be determined (Struijs et al., 2014), the emerging paradigm of Big data, involving several domains in the public and private sectors, warrants critical scrutiny (Crawford, et al., 2014).

Big Data as Cultural and Corporate Capital

When it comes to defining Big data, the convoluted participation of actors with very dissimilar philosophies has complicated its conceptualisation. Puschmann & Burgess (2014) explain that journalists, company executives, lawmakers, and academics often employ –subconsciously or strategically– metaphoric terminology to shape Big data's abstractness. The two predominant metaphors frequently used are

largely powered by market-driven logics. On the one hand, "Big data is a force of nature to be controlled" (Puschmann & Burgess, 2014, p. 1699) and on the other "Big data is nourishment/fuel to be consumed" (Puschmann & Burgess, 2014, p. 1700). This neoliberal attitude towards Big data frames it as a natural resource, like the water that is "all at once essential, valuable, difficult to control, and ubiquitous" (Puschmann & Burgess, 2014, p. 1699), but which also threatens us with "torrents of data in which one can drown, floods that overwhelm us, and tsunamis that leave destruction in their wake" (Puschmann & Burgess, 2014, p. 1699). Nonetheless, these views have failed to recognise –intentionally or not– that data is created by organisations and citizens. That it is not merely "a naturally occurring resource of universal cultural relevance that exists without human intervention, [whose] potential can be harnessed through the use of appropriate technology (dams, irrigation)" (Puschmann & Burgess, 2014, p. 1699).

Scholarly research that engages more critically with notions of Big data recognises that data forms the basis of information and knowledge, as it is the product of "abstracting the world into categories, measures and other representational forms" (Kitchin, 2014, p. 1). In this process of measuring, representing, categorising, quantifying and making sense of the world where we live (Crawford, et al., 2014; Mayer-Schonberger & Cukier, 2013; Kitchin, 2014) data has "strong utility and high value because [it] provide[s] the key inputs to the various modes of analysis that individuals, institutions, businesses and science employ in this process (Kitchin, 2014, p. 1). Yet a distinction needs to be made, since databases and data infrastructures do not function as impartial apparatuses for the management of data. Rather, they work as "complex sociotechnical systems that are embedded within a larger institutional landscape of researchers, institutions and corporations, constituting essential tools in the production of knowledge, governance and capital" (Kitchin, 2014, p. 21).

Whilst human interaction becomes increasingly mediated by data with a growing capacity to predict social phenomena, it simultaneously becomes an inescapable instrument of control (Puschmann & Burgess, 2014 p. 1691; Crawford, et al., 2014). Rodley & Burrell, for instance, approach data as a tool for "thinking about and critiquing the hegemonic forces that control, monitor, and police the internet's information portals" (2014, p. 85) as the data surveillance performed by corporations and governments ensures a tight control over the data made available to citizens. O'Reilly (2013) remarks that these interactions between citizens, businesses and government are ultimately subject to algorithmic regulation, that is, the use of algorithms and automations to monitor and police the activities of individuals and organisations. It is unclear, however, who regulates and ensures that these algorithms are performing their roles effectively.

Despite of the existence of data protection laws, largely ineffective nowadays as individuals willingly share data online, Mayer-Schonberger & Cukier suggest that

individuals [will] shift from privacy to probability: algorithms will predict the likelihood that one will get a heart attack (and pay more for health insurance), default on a mortgage (and be denied a loan), or commit a crime (and perhaps get arrested in advance). It leads to an ethical consideration of the role of free will versus the dictatorship of data. (2013, p. 17)

It is precisely the predominance and uncertainty surrounding Big data that calls for a renewed ontology of data literacy. A new ontology of data literacy must necessarily recognise that every stage of data processing and sense-making can benefit exponentially from not only the introduction of computational tools and automations, but the adoption of computational thinking (Wing, 2006, cited in Gynnild, 2013)

by professionals and citizens alike. Computational thinking refers to an aspect of human cognition that draws on concepts from computer science and seeks to understand human behaviour through a problem-solving and system-design mentality (Gynnild, 2013). As the addition of data scientists becomes mainstream in every institution of society producing or handling data, open source philosophies, such as data transparency (Aitamurto, et al., 2011; Kitchin, 2014) are increasingly prevalent in the work routines of such organisations and bodies.

Considering the example of news media again, Parasie and Dagiral (2012) and Felle (2013) agree that a computational data journalism approach can arguably contribute more effectively to democracy by a) accurately analysing data, thus strengthening journalistic objectivity; b) designing and developing inexpensive tools to sustain government accountability; and c) enhancing public participation for the production and analysis of data. Nonetheless, boyd & Crawford display a degree of scepticism towards claims of data objectivity and accuracy. In this respect, the authors believe that data is subject to limitation and bias, and a failure to acknowledge their existence and understand the implications of these variables leads to misrepresentation (Boyd & Crawford, 2012). They further remark that the value of data analysis is undeniable, "yet retaining context remains critical, particularly for certain lines of inquiry" (boyd & Crawford, 2012, p. 670), and this is where journalists, educators, historians, and librarians can make a solid contribution.

Nowadays, computing software simulation "liberates media creation and interaction techniques from their respective hardware" (Manovich, 2013, p. 199-200). By translating these techniques into separate algorithms, software simulation opens up a range of possibilities for data literacy, but only if citizens and professionals engaging with data are fully conversant with databases and algorithms, and are also well aware of their cultural impact.

Data Literacy and Civic Engagement

Thus far, it can be argued that a mismatch between the data competences of professionals working for data brokers/holders and the data literacy of citizens is fairly evident. It is not only that numeracy, statistical literacy or data literacy appears to be an exclusive skillset of mathematicians, statisticians, social, computing or data scientists, but also that a broader and more critical understanding of data as a cultural artefact mostly mediated by corporate and financial interests, appears generally limited for citizens overall. Indeed, a recent report by IBM states that "data-rich, analytically-driven enterprises are 2.7 times more likely to educate employees on the use of data and analytics […] raising the data literacy rate" (IBM, 2014). This statement not only clearly emphasises the predominance of corporate ideologies on data literacy dynamics and schemes, but also evidences how access to ways of handling and making sense of data are generally restricted to those individuals working for data brokers.

Within this context, Mandinach and Gummer (2013) define data literacy as

The ability to understand and use data effectively to inform decisions. It is composed of a specific skill set and knowledge base that enables educators to transform data into information and ultimately into actionable knowledge (Mandinach, Honey, Light, & Brunner, 2008). These skills include knowing how to identify, collect, organize, analyze, summarize, and prioritize data. They also include how to develop hypotheses, identify problems, interpret the data, and determine, plan, implement, and monitor courses of action. The decisions that educators need to use data to inform are multiple and diverse, and data

literacy is tailored to the specific use. For instance, teachers need to combine data literacy with pedagogical content knowledge (Shulman, 1986) to affect instructional practice. Mandinach (2009, 2012) has termed this pedagogical data literacy, whereas Means, Chen, DeBarger, and Padilla (2011) refer to it as instructional decision making. (2013, p. 30)

Mandinach and Gummer's definition is essential for our purposes, as it calls for the inclusion of data literacy ideals into the pedagogical content of subjects such as Science and Mathematics. Furthermore, Schield (2004) sees information literacy, statistical literacy and data literacy as essential and interrelated instruments for the effective evaluation of information, and the predominance of one or another will depend on which discipline undertakes such evaluation in the first place.

Following up on Mandinach & Gummer's remarks, the permeation of data literacy into the teaching and learning of Mathematics and Science appears essential to influence instructional practice. In this respect, they explain that policymakers require educators to make sense of data to then inform practice (2013), which ultimately becomes a key factor in the consolidation of a self-sustainable educational system encouraging civic data literacy. Although they stress the hesitancy of educators to enhance their data literary skills, and their wariness towards technologies for data sorting, cleaning and processing, Mandinach & Gummer conclude that a) "data-driven decision making must become part of an educator's preparation" and b) "Schools of education must find ways to integrate data-driven practices and principles into the training of educators" (2013, p. 34).

CHALLENGES TO THE ADOPTION OF BIG DATA LOGICS

So far, I have attempted to outline the argument that data literacy, regardless of the subject through which it is taught, forms the foundations of our current understanding of societal data-driven dynamics and institutions. Secondly, whilst the institutionalisation of knowledge on Big data primarily occurs within datafied corporations, educators should assimilate data logics and schools should promote datafication as a means to democratise data literacy. Finally, civic society in general needs to engage not only with the instrumental aspects of data, but more importantly, with its critical appraisal as a cultural form that mediates power. In order to achieve these three ideals, I will propose a series of four key challenges to overcome.

Challenge 1: Understanding Data as a Material Force

Digital data is often assumed to be a by-product of the virtual world. Binary information interpreted by machines, which in turn format such information to be readable by humans. And yet digital data has begun to contradict its virtuality in favour of a more tangible materiality, ever since the experimental verification of Landauer's principle by Bérut et al (2012). This principle linked information and thermodynamics, and facilitated the empirical determination of the amount of energy needed to delete a bit of computation.

In its broadest sense, data may be understood as either numerical information contained in tables, or raw data allocated through databases or datasets. Nonetheless, here I want to establish a distinction between two dimensions of data materiality. On the one hand, data might act as evidential facts that inform decisions at government or corporate level. On the other hand, data might work as computational

information powering digital objects used in everyday life. Therefore, in my attempt to expand beyond the reductionist characterisation of data as evidence; and hence controllable, inactive and subordinate to human agency, I characterise it through new materialist ontologies as 'something more than "mere" matter: an excess, force, vitality, relationality, or difference that renders matter active, self-creative, productive, unpredictable' (Coole & Frost, 2010, p. 9).

So by shifting away from a 'vulgar' view of data as a mere artefact, I focus on 'a theory that claims to entirely transcend the dualism of subjects and objects' (Miller, 2005, p. 3) leaving behind the typical tool-orientation of objectifying research towards an understanding of 1) how data is mediated by the constraints of its own materiality?—the limited resources of processing power, storage, and connectivity (Blanchette, 2011); and 2) how it mediates authority, hierarchy, power, routines, and practices by standing in as the actor playing the part of social structure? (Neff, 2015). Within this mindset, data powers algorithms to function as bodies that are relatively independent of human authorisation. This partial autonomy from human use or authorisation enables automated systems to embody what Deleuze and Guattari term a 'body without organs' (1987, p. 4); that is, hybrid entities that are 'neither entirely material, nor are they entirely human' (Anderson, 2013, p. 1016) and that are capable of making decisions based on rules articulated by programmers or through machine-learning (Diakopoulos, 2015).

Understanding data – through its more active embodiment, algorithms – thus requires ever increasing rates of data literacy. As individuals actively engage with data, it pervades their reflexivity, enhancing their computational cognition (Pylyshyn, 1986; Rapaport, 1995). Note that here I make no distinction between the *reflexivity* of both data professionals and the intervening automations involved in the production of data outputs, because in my view they display similar mind [and system] processes, similar biases, and perform decision-making following similar patterns and stages. In fact, by approaching data as a material force similar to that of evolution, we can acknowledge that data and algorithms are strong determinants of our destiny as human beings. For instance, web metrics have become a powerful aid for news editors to decide what becomes news. Algorithms measure which news stories are most liked by the public and that insight is used in editorial decision making to create the news agenda. It is clear then that matters of public concern discussed every day, are nowadays strongly mediated by automated technologies unknown to the public. Hence, it is important to understand the instrumental dynamics of data; namely the statistical operations that govern algorithmic decision-making, or the irrelevance of sampling for Big data processing. However, by acknowledging data materiality beyond objectifying views, we can better understand the role of embryonic forms of artificial intelligence in the production of knowledge, the negotiation of societal power, the parameterisation of the world and social life, the potential remediation of humans, and algorithmic accountability.

Challenge 2: Understanding Data as the Vital Fluid of Societal Institutions

Big data is unquestionably one of the key areas for productive development as governments and industry alike face constant challenges in relation to the management of Big data on the one hand, and exploring creative and innovative ways of engaging with data for the development of society, on the other. The defining features of media nowadays are strongly negotiated by computing software. As suggested by Manovich, media can be characterised by the conjunction of algorithms and data structures (2013). Interestingly, data mediation is not restricted solely to media organisations. Virtually every sector of society currently deals with both media and computing software on a regular basis. A rising institutional dependence on database architecture and algorithmic automation ensures that society's machinery is well

oiled and working smoothly. Consequently, in a move that has gone unnoticed by those responsible for holding the "powers that be" accountable, control is shifting from traditional spheres to organisations which nourish the wealth of data.

Here I argue that in order to understand how profoundly data is embedded in the structures of public bodies, governments, corporations, and civic society overall, educators must adopt a more correlational approach typical of computing experts. These attitudes towards discovering patterns and correlations in the data to unearth novel and invaluable insights are complementary to an understanding of the notions of media numerical representation, modularity, automation, variability and transcoding (Manovich, 2001) that are essential for recognising how effectively data governs the quantification of the world. Therefore, I envisage two conflicting scenarios. On the one hand education has to operate in increasingly datafied schools where the growing predominance of a computerised paradigm is evident – as is demonstrated by the increasing pressure to implement educational programmes on computing coding, for instance. On the other hand, educators have to confront – and try to palliate through the enhancement of civic data literacy – the hegemony of an increasingly powerful *datocratic* elite, which drives policymaking and corporate growth.

As long as a significant portion of educators is bound to data-sceptical paradigms, they remain unaware of the fundamental ways in which Big data, computing logics, database systems and algorithms are transforming modern cultures and societies. Williamson in his critique to ClassDojo (the fastest growing educational technology in history) explains:

Like many emerging ed-tech products, it is also the product of a feverish rush by venture capital firms and philanthropic individuals to invest in education. It represents a growing symmetry between psychological innovation and technical innovation in Silicon Valley. By encouraging children to see themselves as malleable, adaptable and improvable, it treats children in exactly the same way that high-tech companies from the valley treat their employees. (2016).

As conventional data-sceptical paradigms appear disconnected from the fast-paced, database-driven contemporary world, an urgent necessity for educational institutions able to foster data literacy in order to confront data in its own terrain is becoming imperative. Only by shifting from outdated paradigms to critically engage with computational ontologies that facilitate the understanding of data in a broader sense, will educators make a valuable contribution to the proliferation of datafied civic engagement and the regulation of data systems, algorithms and data organisations in our *datocratic* society. Although educators increasingly show a burgeoning awareness of elements of control and power within data structures, organisations and actors (Selwyn, 2015), additional computational enablers are required to reach a more sophisticated understanding of contemporary *datocratic* dynamics.

Challenge 3: Understanding Notions of Data Access and Data Generation

As scholars continue exploring the complexities of Big data, this third challenge underlines the tension between two key data literacy approaches: a dominant one, whereby data is approached following axiomatic norms and instrumental rules based on elementary numeracy and statistics; and an emerging

one, whereby the use of methods and logics from data science enable educators to interact with data in innovative ways. Drawing on this initial distinction, this section explores notions of data access and data generation, arguing that a healthy and wide-ranging data literacy will remain unreachable as long as the traditional approach of using numbers as hard facts, and computers as tools for data processing, persists.

In order to facilitate the transition from one approach to the other, it remains essential to identify the main data holders and data gatherers – responsible for granting not only access to the data but also to the technologies that process it. At present, Big data is seemingly mediated and processed by different types of institutions with different kinds of agendas. Here I am particularly interested in exploring the gatekeeper function (Lewin, 1947) and those responsible within public sector organisations for making open data available, particularly in relation to the Freedom of Information Act. The core assumption regarding gatekeepers is that these individuals decide what information, which in the context of this study is a commodity, is made accessible and what is not (McCombs, 1978). Gatekeepers are most often viewed as individuals but can also be institutions or organisations that manage, facilitate, and restrict the flow of information and therefore have power and the ability to influence citizens' empowerment and the construction of their social reality (McCombs, 1978). Identification of gatekeepers is helpful in that it provides "useful analytic devices for learning about the vocabularies of structure in an organization" (Morrill, et al., 1999, p.51). Within public sector organisations, formal gatekeepers have significant power, as they are able to grant or deny access to information (such as labelling a request vexatious). Furthermore, the degree of citizen access to big and open data held by governments seems to be small in certain contexts (García & Gertrudix, 2011). For all these reasons, it is of paramount importance to understand the people who perform this mediation, the institutions within which they function, the principles and policies that guide their work and the impact of these processes and interactions on citizen empowerment.

The varying ways in which data is managed by individuals or institutions seem to suggest that data access is connected to –if not the key determinant of– the way these actors perceive data value. Common practice indicates that social actors normally engage with four types of data: open data supplied by public institutions following transparency and openness schemes; public data obtained through negotiation or formal FOI requests, proprietary data made available under paid subscription by companies who either own the rights for commercialisation or produce the data from scratch; and data that is protected behind a veil of secrecy and to which privileged actors only have access through whistleblowers or leaks.

So, to summarise this challenge, I reflect on the diverse power dynamics negotiated through data access and the perceived value of its acquisition. As the ability to access certain data structures influences data's perceived value, societal institutions, be they public or private, have historically commoditised proprietary data, overly emphasising the monetary value of outputs derived from it. Paradoxically, freely available data accessed through both open data regimes and FOI requests is generally treated with misgiving, and has a lower perceived value, as arguably it is provided by press offices and spin doctors following political agendas. Finally, and more importantly, the most valuable data is that which is withheld by corporations, and that cannot be accessed through traditional routes. This argument raises an interesting dichotomy whereby access to data can be mediated through the use of data technology (API wrangling, reverse engineering, web scraping), as long as individuals or institutions are not dealing with inaccessible data, in which case less conventional methods of acquisition –such as leaks– would be necessary.

Challenge 4: Understanding the Differences Between Public and corporate Data and How They Mediate Power

Today, emerging technologies pose new challenges as society battles to keep up with the pace in a fast-changing world, particularly in the light of the advent of Big data, which has had ramifications at various levels of society and has become a key area for government, and economic growth. In a previous research paper looking at how British news media constructed the notion of Big data, I concluded that deliberation on this topic primarily occurred within the financial and corporate spheres (particularly framed by innovation discourses) with coverage on Big data being predominantly placed within Business sections, and framed with an economically optimistic tone as a generator of growth (Borges-Rey, 2015). My analysis also suggested that much of this business coverage was driven by external consultants and other outsourced actors who influenced the discourse towards a positive evaluation. Complementary critical discourse analysis suggested that argumentations were normally framed as advice and guidance to CEOs and managers on how to harness the power of Big data to increase the profits of their companies (Borges-Rey, 2015).

As the effectiveness of data organisations in monetising the insights derived from citizens' data increases, so does the power they hold over not only the individuals, but also over the institutions of society. Corporations such as Google and Facebook, with a core focus on quantifying everyday life, have coded algorithms capable of profiling and predicting people's hopes and dreams in an environment free of public or institutional scrutiny. In the past, this watchdog function was performed by news media as part of a healthy democratic society. Nowadays, news organisations seem to be unable to monitor the contemporary institutional negotiation of data power, as it arises in a scenario only accessible to actors with a competent degree of computational thinking or cognition.

So in this final section I will argue that there is an imbalance in the rhetoric, wherein Big data is predominantly framed as the epicentre of contemporary innovation and the driving force of societal progress. This prevalent and tacitly neoliberal discourse appears to safeguard corporate data commodities to the detriment of overly exploited public data, making the data generated within public bodies and by citizens a profitable good to nourish. This dynamic emphasises a mayor gap between the data brokers who, through a series of furtive transactions, exploit and commoditise citizens' data, and citizens who display varying degrees of awareness regarding the amount of data they generate on social media or that is generated by public bodies, and how that data is utilised by increasingly powerful data brokers. Notably, a number of civic-driven initiatives operating in platforms such as GitHub or Hacks/Hackers, and start-ups such as ScraperWiki, are now essential spaces and proficient enablers of data-driven public deliberation and civic engagement, thus fostering an ever-expanding rate of data literacy and improving public participation.

CONCLUSION

This chapter posited a series of challenges that arise when Big data is not properly contextualised in the delivery of educational strategies for the enhancement of data literacy. In this vein, this research considered the materiality of computerised data in order to examine its implications for data literacy. It also examined the ways that notions of data access, data sampling, data sense-making and data collection

are nowadays intermediated or contested by datafied governments and corporations, thus hindering the capacity of citizens to effectively understand and make better use of the data they generate or engage with.

Firstly, I made an argument for understanding data as a material force. That is, by assuming data's materiality as a driving force for the evolution of datafied dynamics, institutions, individuals and cultures. This proposition requires understanding data beyond objectifying views that generally regarded data as evidential information to inform decisions at various levels in society. By recognising data as the material force that powers algorithms (bodies without organs); automations and other forms of artificial intelligence act as peers in pedagogical strategies, freeing educators up from the tedious low-level work of processing, sorting, and cleaning data, so that they can focus on enhancing data literacy, or informing policymaking and instructional decision-making. Data as a material force also helps to understand the complex power dynamics that occur in society and greatly impact the generation of knowledge. Here I drew a distinction between the two ontologies of data materiality mediating the main approaches to Big data in education. On the one hand data acts as evidential hard facts that inform decision-making; on the other hand, data powers embryonic forms of artificial intelligence that aid educators in their search for better and more widespread data literacy. This second approach arguably enables an awareness of not only the importance of untangling the complexities of a world driven by data and algorithmic automation, but more importantly the role of Science and Mathematics educators as transcoders in a society that is rather illiterate in aspects concerning data.

Secondly, I suggested the prevalence of a data-sceptical paradigm amongst educators, deeply rooted in an objectifying mindset. As long as educators are bound to this traditional paradigm, they remain unaware of the profound ways in which Big data, computing logics, database systems and algorithms are transforming institutional cultures in terms of privacy, data literacy, automation, interactivity, virtuality, and engagement. This objectifying approach therefore seems disconnected from the fast-paced, database-driven contemporary world, and an urgent necessity to shift towards a critical appraisal of computational approaches and to understand data in a broader sense seems imperative. By engaging with datafied logics in such terms, educators are more likely to make a valuable contribution to the proliferation of data-driven civic engagement and the regulation of data systems, algorithms and data organisations.

Thirdly, I argued that in order to facilitate the transition from one approach to the other, it remains critical to identify the main data holders and data gatherers – those responsible for granting not only access to the data but also to the technologies to process it. A general inability to access data beyond the usual channels (open data platforms, freedom of information regimes, or similar) appears to be linked to a lack of computing skills. Although educators in the fields of Mathematics and Science are generally well equipped to undertake instrumental processing of data – displaying a highly competent degree of information, statistical or data literacy – data infrastructures nowadays appear to also necessitate a degree of computational thinking or cognition. As explained by boyd & Crawford, practices such as "Wrangling APIs, scraping, and analysing big swathes of data are generally restricted to those with a computational background" (2012, p. 673) and educators with no access to those skills are perceived as being disadvantaged in the hierarchies around "who can read the numbers" (Boyd & Crawford, 2012, p. 673). In this respect, analysing an extensive database manually or relying on data disclosed by public organisations appears to be a setback. This is particularly true when the development of AI and Semantic Web technologies is so advanced that it can bring real value to educational institutions by curating web-based content, and creating the tools to measure the accuracy, timeliness, and comprehensiveness of informational data.

Finally, I have argued that a change of mindset appears essential. Numbers are no longer the raw material that forms evidential accounts nor are they objective measurements used to make sense of the world and its social problems. In the age of Big data, numbers are the codes and frameworks that manage innovative outputs and strategies that fuel an ideal for widespread data literacy. Numbers are also the parameters and variables that govern the algorithms and computing systems that regulate our everyday life; and as such, need to be interrogated and held accountable. Finally, numbers are the inputs that data organisations employ in the datafication (Mayer-Schonberger & Cukier, 2013) of social interaction, geography and knowledge. Only by shifting from an objectifying paradigm to a critically-computerised one will educators and their pupils understand data in a broader sense. As a new generation of early adopters accustomed to consuming multilinear object-oriented narrative forms (Pavlik, 2001) through database browsing techniques moves towards the demographics targeted by several data-brokers, a burgeoning awareness of elements of control and power within data structures, organisations and actors is noticeable within these collectives. Nevertheless, additional computational enablers are required to reach a more sophisticated understanding of contemporary *datocratic* dynamics.

REFERENCES

Aitamurto, T., Sirkkunen, E., & Lehtonen, P. (2011). *Trends in Data Journalism.* Next Media, a Tivit programme.

Anderson, C. W. (2013). Towards a sociology of computational and algorithmic journalism. *New Media & Society*, *15*(7), 1005–1021.

Bérut, A., Arakelyan, A., Petrosyan, A., Ciliberto, S., Dillenschneider, R., & Lutz, E. (2012). Experimental verification of Landauer/'s principle linking information and thermodynamics. *Nature*, *483*(7388), 187–189. doi:10.1038/nature10872 PMID:22398556

Best, J. (2012). *Damned lies and statistics: untangling numbers from the media, politicians, and activists*. London: University of California Press.

Blanchette, J. F. (2011). A material history of bits. *Journal of the American Society for Information Science and Technology*, *62*(6), 1042–1057. doi:10.1002/asi.21542

Borges-Rey, E. (2015). Framing discourse on big data: Online coverage of the big data revolution by British newspapers.*Data Power Conference*. University of Sheffield.

boyd, d., & Crawford, K. (2012). Critical questions for big data. Information Communication and Society, 15(5), 662–679.

Chevallard, Y. (2013). Enseñar matemáticas en la sociedad de mañana: Alegato a favor de un contraparadigma emergente. *Journal for Research in Mathematics Education*, *2*(2), 161–182.

Cohn, V., & Cope, L. (2012). *News & numbers* (3rd ed.). Sussex, UK: Wiley-Blackwell.

Coole, D., & Frost, S. (2010). Introducing the New Materialisms. In D. Coole & S. Frost (Eds.), *New materialisms: Ontology, agency, and politics* (pp. 1–43). Duke University Press. doi:10.1215/9780822392996-001

Crawford, K., Miltner, K., & Gray, M. L. (2014). Critiquing Big Data: Politics, Ethics, Epistemology. *International Journal of Communication*, *8*, 1663–1672.

Deleuze, G., & Guattari, F. (1987). *A thousand plateaus*. London: University of Minnesota.

Diakopoulos, N. (2015). Algorithmic Accountability. *Digital Journalism*, *3*(3), 398–415. doi:10.1080/21670811.2014.976411

Felle, T. (2013). Old reporting, new methods: how data journalism is keeping an eye on government. In J. Mair & R. L. Keeble (Eds.), *Data journalism: Mapping the future* (pp. 133–147). Suffolk, UK: Abramis.

García, F., & Gertrudix, M. (2011). Naturaleza y características de los servicios y los contenidos digitales abiertos. *CIC Cuadernos de Información y Comunicación*, *16*, 115–138.

Gynnild, A. (2013). Journalism innovation leads to innovation journalism: The impact of computational exploration on changing mindsets. *Journalism*, *15*(6), 713–730. doi:10.1177/1464884913486393

Hayles, K. N. (2012). *How we think: digital media and contemporary technogenesis*. London: The University of Chicago Press. doi:10.7208/chicago/9780226321370.001.0001

Howard, A. (2012). *Data for the public good*. O'Reilly.

Huff, D. (1973). *How to lie with statistics*. London: Penguin Books.

Hunt, K. (2004). The challenges of integrating data literacy into the curriculum in an undergraduate institution. *IASSIST Quarterly* (Summer/Fall).

IBM Centre for Applied Insights. (2014). *Inside the mind of Generation D: What it means to be data-rich and analytically driven*. Retrieved from: http://www.ibm.com/smarterplanet/us/en/centerforappliedinsights/article/gen_d_insights.html

Jacobs, J., Gregory, A., Hoppey, D., & Yendol-Hoppey, D. (2009). Data literacy: Understanding teachers data use in a context of accountability and response to intervention. *Action in Teacher Education*, *31*(3), 41–55. doi:10.1080/01626620.2009.10463527

Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. London: Sage. doi:10.4135/9781473909472

Lewin, K. (1947). Group decision and social change. *Readings in Social Psychology, 3*, 197-211.

Maier, S. R. (2002). Numbers in the News: A mathematics audit of a daily newspaper. *Journalism Studies*, *3*(4), 507–519. doi:10.1080/1461670022000019191

Mandinach, E. B., & Gummer, E. S. (2013). A systemic view of implementing data literacy in educator preparation. *Educational Researcher*, *42*(1), 30–37. doi:10.3102/0013189X12459803

Manovich, L. (2001). *The language of new media*. MIT Press.

Manovich, L. (2013). *Software takes command.* London: Bloomsbury Academic.

Mattelart, A. (2005). *The information society: an introduction*. London: Sage.

Mayer-Schonberger, V., & Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work and Think*. London: John Murray.

McCombs, M. (1978). Agenda setting function of mass media. *Public Relations Review*, *3*(4), 89–95. doi:10.1016/S0363-8111(77)80008-8

Meyer, R. (2014, June 28). Everything We Know About Facebook's Secret Mood Manipulation Experiment. *The Atlantic*. Retrieved from http://www.theatlantic.com/technology/archive/2014/06/everything-we-know-about-facebooks-secret-mood-manipulation-experiment/373648/

Miller, D. (2005). Materiality: an introduction. In D. Miller (Ed.), *Materiality* (pp. 1–50). Duke University Press. doi:10.1215/9780822386711-001

Moore, D. W. (2009). *The opinion makers: An insider exposes the truth behind the polls*. Boston: Beacon Press.

Morrill, C., Buller, D. B., Buller, M. K., & Larkey, L. L. (1999). Toward an organizational perspective on identifying and managing formal gatekeepers. *Qualitative Sociology*, *22*(1), 51–72. doi:10.1023/A:1022183232593

Neff, G. (2015). Learning from documents: Applying new theories of materiality to journalism. *Journalism*, *16*(1), 74–78. doi:10.1177/1464884914549294

O'Reilly, T. (2013). Open data and algorithmic regulation. In B. Goldstein & L. Dyson (Eds.), *Beyond transparency: open data and the future of civic innovation* (pp. 289–300). San Francisco: Code for America Press.

Parasie, S., & Dagiral, E. (2013). Data-driven journalism and the public good: Computer-assisted-reporters and programmer-journalists in Chicago. *New Media & Society*, *15*(6), 853–871. doi:10.1177/1461444812463345

Pavlik, J. V. (2001). *Journalism and new media*. New York: Columbia University Press. doi:10.7312/pavl11482

Puschmann, C., & Burgess, J. (2014). Metaphors of Big Data. *International Journal of Communication*, *8*, 1690–1709.

Qin, J., & D'Ignazio, J. (2010). *Lessons learned from a two-year experience in science data literacy education.* Academic Press.

Rapaport, W. J. (1995). Understanding understanding: Syntactic semantics and computational cognition. *Philosophical Perspectives*, *9*, 49–88. doi:10.2307/2214212

Rodley, C., & Burrell, A. (2014). On the art of writing with data. In J. Potts (Ed.), *The future of writing* (pp. 77–89). Basingstoke, UK: Palgrave Macmillan. doi:10.1057/9781137440402_7

Schield, M. (2004). Information literacy, statistical literacy and data literacy. *IASSIST Quarterly*, *28*(2/3), 6–11.

Selwyn, N. (2015). Data entry: Towards the critical study of digital data and education. *Learning, Media and Technology*, *40*(1), 64–82. doi:10.1080/17439884.2014.921628

Stephenson, E., & Schifter Caravello, P. (2007). Incorporating data literacy into undergraduate information literacy programs in the social sciences: A pilot project. *RSR. Reference Services Review*, *35*(4), 525–540. doi:10.1108/00907320710838354

Struijs, P., Braaksma, B., & Daas, P. (2014). Official statistics and Big Data. *Big Data & Society*. Retrieved from: http://bds.sagepub.com/content/1/1/2053951714538417.abstract

Treanor, J. (2015, April 22). The 2010 'flash crash': how it unfolded. *The Guardian.* Retrieved from https://www.theguardian.com/business/2015/apr/22/2010-flash-crash-new-york-stock-exchange-unfolded

Williamson, B. (2016, February 16). *Decoding ClassDojo* [Blog post]. Retrieved from https://codeactsineducation.wordpress.com/2016/02/16/decoding-classdojo/